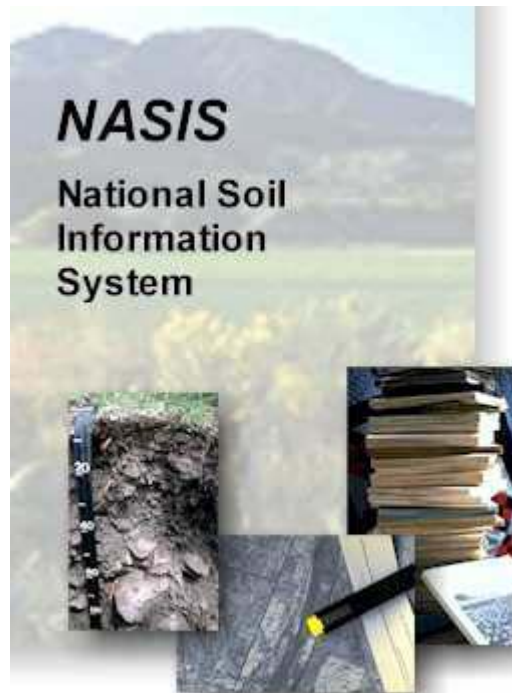


Soil Data Delivery And Distribution



Interim Report, June 18, 2002
Total Requirements Statement

Acknowledgments

This is a preliminary release of the Total Requirements Statement for Soil Data Delivery and Distribution. It has not been through a management review and approval. Information and details for specific business requirements are subject to change.

This report is the result of the cumulative effort of the following principal contributors.

USDA - Natural Resources Conservation Service
Soil Survey Division
Washington, D.C.

Information Technology Center
Fort Collins, Colorado

Executive Sponsor: Berman Hudson, Director, Soil Survey Division
Sponsor Representative: Russ Kelsea, National Leader, Soil Survey Technical Services
Project Manager: Ken Harward
Application Analyst: Gary Spivak
System Analyst: Dorn Egley
Business Analysts:

- | | |
|---------------------------------------|---|
| <input type="checkbox"/> Terry L. Aho | <input type="checkbox"/> Rick Bigler |
| <input type="checkbox"/> Jim Fortner | <input type="checkbox"/> Jennifer Sweet |
| <input type="checkbox"/> Ken Lubich | <input type="checkbox"/> Fred Minzenmayer |

Additional help was received from a number of subject matter experts and end users of soils information who assisted in determining the requirements for soil data delivery and distribution.

DISCLAIMER: Trade names are used solely to provide specific information. Mention of a trade name does not constitute a guarantee of the products by the U.S. Department of Agriculture nor does it imply endorsement by the Department or the Natural Resources Conservation Service over comparable products that are not named.

The U.S. Department of Agriculture (USDA) prohibits discrimination in all its programs and activities on the basis of race, color, national origin, gender, religion, age, disability, political beliefs, sexual orientation, and marital or family status. (Not all prohibited bases apply to all programs.) Persons with disabilities who require alternative means for communication of program information (Braille, large print, audiotape, etc.) should contact USDA's TARGET Center at 202-720-2600 (voice and TDD).

To file a complaint of discrimination, write USDA, Director, Office of Civil Rights, Room 326W, Whitten Building, 14th and Independence Avenue, SW, Washington, DC 20250-9410 or call (202) 720-5964 (voice or TDD). USDA is an equal opportunity provider and employer.

Table of Contents

1	Perspective	1
2	Warehouse Database and Staging Server	5
2.1	<i>Export Tabular Data from NASIS to Staging Server</i>	5
2.2	<i>Export Soil Supporting Data (Later Phase)</i>	6
2.3	<i>Export Soil Spatial Data to Staging Server</i>	7
2.4	<i>Import Data into Staging Server (D3)</i>	8
2.5	<i>Evaluate and Approve Quality of Data and Release to Warehouse</i>	12
2.6	<i>Warehouse Database (D5)</i>	15
2.7	<i>Export Soil Data to FOTG and SSURGO Data Mart</i>	17
3	FOTG and SSURGO Data Mart	18
3.1	<i>Catalog of Available Data</i>	18
3.2	<i>Select Data</i>	18
3.3	<i>Order Data</i>	19
3.4	<i>Register for Notification of Data Updates</i>	20
3.5	<i>Deliver data</i>	20
4	Data Flow Diagrams	21
4.1	<i>Legend</i>	21
4.2	<i>Importing Soil Survey Data into the Warehouse</i>	22
4.3	<i>Importing Laboratory Data into the Warehouse</i>	23
4.4	<i>FOTG and SSURGO Data Mart</i>	24
5	Data Model (Not Complete)	25
5.1	<i>Business Rules</i>	25
5.2	<i>Tables and Attributes (Not Complete)</i>	27
6	Web Soil Data Viewer Data Mart (Not Complete)	28
6.1	<i>Navigate</i>	28
6.2	<i>Show data availability scale dependency (STATSGO/SSURGO)</i>	29
6.3	<i>Web Soil Data Viewer Interface</i>	29
7	Soil Reports and Interpretations Data Mart (Not Complete)	30
7.1	<i>Select Data</i>	30
7.2	<i>Register for Notification of Data Updates</i>	32
8	Custom Export Data Mart (Not Complete)	33
8.1	<i>Web Access Interface</i>	33

8.2	<i>Catalog of Available Data</i>	33
8.3	<i>Select Data</i>	34
8.4	<i>Order Data</i>	35
8.5	<i>Register for Notification of Data Updates</i>	35
8.6	<i>Deliver data</i>	36
9	Application Programming Interface (API) Data Mart (Not Complete)	37
10	Integration with Other Resource Databases (Not Complete)	38
Appendix A	Fundamental Requirements	41
Appendix B	Priority List of Fundamental Requirements	51
Appendix C	Design Unit Descriptions	55

List of Figures

Figure 1: Legend of symbols used in Data Flow Diagrams (DFD).....	21
Figure 2: Data Flow for importing soil data into warehouse and delivering data to FOTG Data Mart.....	22
Figure 3: Soil Characterization Data Flow from Laboratory Information Management System (LIMS) to Warehouse.....	23
Figure 4: Data Flow FOTG/SSURGO Data Mart.....	24

1 Perspective

Background

The Total Requirements Statement Phase 1 (TRS) contains detailed description of the high priority business requirements and functionalities for a Soil Data Warehouse to be used in the delivery and distribution of soil survey data.

The Total Requirements Statement is a continuation of analysis of the Draft Requirements Statement (DRS), completed October 2001, which outlined the existing systems, limitations of the existing systems and the fundamental requirements for the delivery and distribution of soil survey data.

The Outline Physical Design (OPD), completed December 2001, provided an overview of the proposed design of a system to meet the requirements described in the DRS. The proposed design is a phased approach. The OPD describes design units, a phased implementation schedule and estimated cost to meet the priority requirements in the DRS.

Objective

Implement an information system to provide for and deliver:

- ❑ A single source of current official soil survey data of high quality that meets USDA national program responsibilities.
- ❑ Access to current and previous versions of official soil survey data.
- ❑ Products that meet customer needs.
- ❑ Customer access for selecting, interpreting, reporting and downloading soil survey data and information.
- ❑ Informational materials explaining the National Cooperative Soil Survey procedures, standards and technical references.
- ❑ Supporting soil survey data such as pedon descriptions, laboratory characterization, and photographic and graphic images.
- ❑ Metadata and other information about the soil survey products.

Management Considerations

Due to budget and resource constraints the implementation and delivery of the information system will be in phases. The first phase targets the high priority business requirements. Additional system capability will be added at a later time as resources become available.

The first implementation phase includes the OPD design units; warehouse database, staging server, and FOTG/SSURGO data mart. Implementing these pieces targets the high priority business requirements, which provide:

- ❑ The secure storage of the current official spatial and tabular soil survey data for an area.
- ❑ Provide stability in delivery format for a defined period of time.

- ❑ Provide versioning and archiving capability and customer access to known versions.
- ❑ Provide customer access to FOTG/SSURGO Data Mart for delivery of official data.

Continued management support is imperative to complete the analysis needed to complete the TRS and the full implementation of the delivery and distribution information system.

Analysis Considerations

The Total Requirements Statement is a continuation of analysis based on the Outline Physical Design (OPD) approved December 2001. The OPD proposed design units that can be implemented in phases and meet groups of business requirements. Design units proposed are:

- ❑ Warehouse Database.
- ❑ Staging Server.
- ❑ FOTG/SSURGO Data Mart.
- ❑ Web Soil Data Viewer Data Mart.
- ❑ Soil Reports and Interpretations Data Mart.
- ❑ Custom Export Data Mart.
- ❑ Soils Application Programming Interface (API) Data Mart.
- ❑ Soil Survey Publication Data Mart

Because of time constraints and uncertain budgets, this phase of the analysis focused on delivery and distribution of official soil survey data. As a result this TRS does not contain a complete analysis of all the business requirements for soil data delivery and distribution.

The first implementation phase includes the design units:

- ❑ Warehouse database.
- ❑ Staging server.
- ❑ FOTG/SSURGO data mart.

The business requirements satisfied by the first phase of implementation are noted in Appendix B, Priority List of Fundamental Requirements.

Analysis needs to continue as resources are available and a second version of the TRS should cover Web Soil Data Viewer Data Mart, Soil Reports and Interpretations Data Mart, Custom Export Data Mart, the Soils Application Programming Interface (API), and importing other data such as Laboratory, interpretations and reports, supporting data, and photographic and graphic images.

Definition of Terms and Symbols

The definitions of terms apply to the usage in this document. Many terms used in this document may have different meanings when used in other contexts.

- ❑ **Certifying Data** – The process conducted by the State Soil Scientist to review and certify soil data for use by field offices and general public.

The state soil scientist can partially certify the data for use for specific applications such as RUSLE, WEPS, Hydric rating, etc. This includes the tabular data and soil maps (digital or analog). This is not intended to represent SSURGO certification process.

- ❑ **Official Data** - The Official Copy of soil survey information is the most current soil information for a soil survey area that is certified for official use within the Field Office Technical Guide by the State Soil Scientist (GM Title 430, Part 402.5). Official data includes the soil map and tabular attribute data.
- ❑ **Tabular Data** – This refers to the attribute data associated with the mapunit data for a soil survey.
- ❑ **Spatial Data** – This refers primarily to the digital soil mapunit that are polygons, points and lines, point and linear special features, and limit of soil survey area.
- ❑ **SSURGO** – Represents the certified SSURGO product (Soil Survey Geographic database) that contains spatial and tabular data and metadata.
- ❑ **Digital spatial data** – Represents a soil survey with spatial and tabular data that may be released to the warehouse that has not gone through the SSURGO certification process.
- ❑ **Most Up-To-Date Data** - Soil data that is most current. It may still be under development and testing (NASIS transactional database) and not officially released for use.
- ❑ **Access** - A user can get the available soil survey data and information. Access does not imply a specific implementation such as web access. Access can include a phone contact to request soil survey information, printed information available at a public source (library, field office, etc), or a web site access.
- ❑ **Soil Survey Supporting Data** - Supporting soil survey data includes site pedon descriptions collected during conduct of a soil survey and technical soil services, soil characterization data, soil performance measurement data, photographic and graphic images, field observation notes and other data collected.
- ❑ **Data Stores** – These are products, models, or systems that are produced or download official soil survey data and information from the soil data origin. They provide that data for use by customers for a period of time separate from the origin. Some data stores are relatively static and the data, such as published soil survey reports, printed interpretative maps and tabular reports can become out dated. Other data stores, such as FOTG, SSURGO and interpretative maps and reports created using Soil Data Viewer are more dynamic and can be updated with more current soil data and information.
- ❑ **Data Marts** – Customer delivery systems that are designed and optimized for the delivery of specific products and services. A data mart can be customized with a specific view of the data, pre-packaged data, or a system that is designed and sized to give optimum performance for expected customer demand.

- ❑ **Special Use Soil Survey Data** – Soil survey data for a given geographic area that are published for a specific customer to meet their needs. These data are not considered as the official soil survey data for use in FOTG. These data might be of different scales, include special interpretations not normally included in the official soil survey data, or advance preliminary data not for official use.

2 Warehouse Database and Staging Server

The warehouse database is the core piece of the system and will be used to provide the single source of official soil survey data authorized for release by the State Soil Scientist. The staging server provides the State Soil Scientist an opportunity to review and validate soil survey data quality, along with interpretation and report functionality before final approval of the data and before the data are placed in the warehouse database for release to public.

The general flow of data is from the transactional databases (NASIS, LIMS, Digitizing Units, etc) to the staging server, and finally to the warehouse database. The sections below outline the requirements of each step of this process.

2.1 Export Tabular Data from NASIS to Staging Server

The primary user is the State Soil Scientist. Soil survey data for an area is selected, reviewed, certified, and exported to the staging server.

2.1.1 Select Data from NASIS transactional database (D1)

NASIS select manager is used to select a legend. Export interface provides the user with the ability to define and select map units, components, text data, and interpretations to include in export.

- Authorized user (State Soil Scientist) selects legend or legends to export.
- Choose map units for legends based on selected set or from permanent tables where map unit status is one or more: provisional, approved, correlated, and/or additional.
- Choose data map units based on those in selected set, all data map units from permanent tables or DMU from permanent tables with certification status of: Not for Distribution, Not Certified, Partly Certified, and/or Certified.
- Choose components based on components in selected set, all components from permanent tables, major components from permanent tables or components from permanent tables with percent composition greater than or equal to XX percent.
- Choose local and/or national interpretations to be generated and included in the export.
- Choose which text kinds from Legend, Map unit, Data Map unit, Component, and Horizon Text tables to include.
- User enters or confirms email notification address: notification occurs when data is loaded into staging server and ready for review and evaluation.
- User initiates export.
- User does not have direct access to export files.

Selected legend not found

- User returns to select manager to retry select query.

- User selects different legend.
- User quits.

2.1.2 Export tabular data from NASIS

The interpretations are generated during the export process. The results of the interpretations are stored for each component.

- Automated process begins loading data into staging server database.
- Email notification is sent to notification address in NASIS export selection interface that data has successfully loaded on staging server.

Export/Import process fails

- Notify user of export problems.
- System cleans up data (temp files) created during export attempt.

2.1.3 Exporting Interpretation and Report Scripts (This is implemented in a later phase)

Authorized user has the capability to select interpretations (rules, evaluations and properties) and report scripts for release to the warehouse. These exported interpretations and reports will be used as part of the Soil Reports and Interpretations Data Mart to be implemented in later phases. Exported interpretations will also be used in amending official soil survey data with the addition of additional interpretations to an already published dataset without risk of updating or changing existing chemical, physical properties, or existing interpretations.

- Authorized user (owner of object) selects interpretations and/or reports.
- Export to staging server is initiated.
- User evaluates and reviews the interpretations and reports on staging server.
- User approves release to warehouse.
- **This is implemented in a later phase. More information needed about export authority in relationship to ownership of interpretations and reports.**

2.2 Export Soil Supporting Data (Later Phase)

Authorized users have the ability to select Site-Pedon and Lab Results for delivery and release to public. The data may come from both NASIS and LIMS. **More detail needed. Who is authorized to export supporting data?**

2.2.1 Authorized Lab user selects the soil lab data from LIMS

Authorized user selects study or project for release to warehouse.

2.2.2 Evaluate and Certify

User evaluates and certifies data complete for release.

2.2.3 NASIS Site-Pedon Data

Authorized user selects Sites and Pedons for export and release to warehouse. Site-Pedon may be associated with characterization data, individual point data or part of transect.

- User can identify specific Pedons for a Site. This case is where more than one version of the Pedon description exists based on corrections and updates. User can choose all Pedons for the site or the most appropriate Pedon for release to warehouse.

2.3 Export Soil Spatial Data to Staging Server

Primary user is Digitizing Unit. The soil spatial data for an area are selected, reviewed, certified and exported.

- State Soil Scientist can select which existing SSURGO datasets represents Official Soil Survey Data for immediate placement in the FOTG/SSUGO data mart.
- The spatial portion of all existing SSURGO will be captured and loaded into the Warehouse. These datasets can be updated with tabular data at a later date.
- SSURGO datasets not identified as official data and not moved to the FOTG/SSURGO data mart and older SSURGO versions 1 are kept publicly available but noted as not official soil survey data (FOTG).

2.3.1 User selects soil spatial data from the digitized soil map database (D2)

- User selects soil survey area for export.
- Metadata is updated with spatial process.
- Existing archive spatial data are extracted from SSURGO archive.

Soil survey area not found

- User tries query again.
- User selects different soil survey area.
- User quits.

2.3.2 User evaluates the completeness, accuracy and certifies spatial data, including metadata

- Spatial review of node connections, labels, survey joins, etc. and that the limit of soil survey area, soil polygons, line and point special features and metadata are without error.
- Needed edits are made.

Data not complete

- User returns to digitizing database and edits and updates data.

- Export process ends.
- User quits.

2.3.3 Spatial data and metadata are transmitted to staging server (D3)

Approved spatial and metadata are sent to staging server.

- User initiates transmittal to staging server.
- Data is bundled and compressed and electronically sent to staging server.

Import fails

- User notified of import problem.
- User edits/updates original soil spatial data.
- User restarts select process.

2.4 Import Data into Staging Server (D3)

The staging server is similar to the NASIS database with the addition of spatial data. The staging server provides for testing and reviewing the data before officially releasing to the public warehouse. Access to staging server is restricted to authorized user for processing soil data to the warehouse.

- User is prompted to overwrite or cancel when importing data for an area that already exists in the staging server.
- For any given geographic area there can be only one set of official soil survey data that represents the same logical entity.
- Geographic reference for the area covered by a specific official soil survey data dataset is a spatial map of area boundary. The initial map is based on the SSURGO area status map.
- Geographic reference for special use soil survey data is a separate spatial map of soil survey area boundaries. This is implemented in a later phase.

2.4.1 Import spatial data

Data from the Digitizing Units are copied to the staging server for verification. This creates a small, temporary database containing any soil surveys that are in the process of verification. They will be removed from the staging server after they are checked into the warehouse.

- Every import must be attributed to either an official area boundary or a special use soil survey area. The area boundary will be checked against existing area boundaries to detect overlaps (see verification process below) and will become part of the official status map or the special use soil survey map. The boundary polygon will be linked to a NASIS Area record.
- Spatial data existing in the Warehouse are extracted when only tabular data for the survey is being updated.
- **NOTE: need more clarification on boundary map, updating, and archive.**

- During the import the digitized features (area and map unit polygons as well as point and line features) are converted to geodatabase format. **Note: More analysis and testing needed with the geodatabase. Concerns of export functionality preserving map accuracy.**
- The digitized polygons are joined with tabular data on the staging server. Spatial datasets that have mukey already populated will be checked and joined. The staging server will allow edit to the keys during the joining process. This can occur where the tabular data from NASIS has a new internal id.
- Attribute tables for special point and line features will also be imported at this point because those data are not stored in NASIS. Unique keys will be assigned. **NOTE: presently keys are only unique within a survey area.**
- That part of the metadata supplied by the DU is also imported at this time and linked to soil survey area.
- Soil maps cannot be imported to the staging server if there is already a map for the same geographic area in the staging server. User is prompted to either overwrite existing data or cancel the import process.

Import digital map failed

- Digitizing Unit is notified.
- DU makes edits and updates Digitizing Unit Database.
- DU re-exports survey area from transactional database.

2.4.2 Import NASIS data

Import NASIS data includes the soil survey tabular data, interpretations (rules, evaluations, properties) and report scripts.

Soil Survey Area Data

- The area symbol is matched with the official area boundary map. If no match is found the area symbol is matched with the special use soil survey area map.
- Data exported from NASIS are loaded into tables in the staging server. This can only be done if records with the same NASIS keys are not already in the staging server. If data already exists the user is prompted to either overwrite existing data or cancel the import process.
- User is notified by Email of successful load and data is ready to evaluate.
- If there are digital soil polygons that need NASIS keys assigned, they are updated with the NASIS map unit key, and linkage records are created.
- If there is not digital soil data the tabular data are loaded for evaluation and approval.

- **Optional implementation:** If the staging server is physically the same as the warehouse database, newly imported objects are flagged as not available for distribution. Object versions already in the warehouse are not imported.

Import legend from NASIS failed

- No area symbol match in either official soil survey area or special use soil survey map.
- State Soil Scientist is notified.
- User evaluates the official area boundary map or special use soil survey area boundary map and updates as needed.
- User evaluates the tabular data in the NASIS transactional database and updates as needed.
- State Soil Scientist re-exports from NASIS.

Interpretation and Report Scripts (Later Phase)

Warehouse interpretations (rules, evaluations and properties) and reports are created in NASIS. They are exported from NASIS and imported into the staging server for final testing and approval. Because of structure and model differences between the NASIS transactional database and the warehouse there is always a possibility that while an interpretation or report works in NASIS it will not work in the warehouse. An authorized user approves release of interpretations and reports to warehouse for public access.

2.4.3 *Import Site-Pedon (Later Phase)*

Importing supporting Site-Pedon data require geographic location or identifier. Site-Pedon data can update or replace existing Site-Pedon. Updating existing data result in versioning. This case includes the delivery of original sampled as data and data that represent observations over time. When replacing existing data the existing data is not versioned. This case is updating with a more current description of the Site-Pedon replacing the existing data that may have errors. **Note: What are the geographic coordinates – Lat/long coordinates?**

2.4.4 *Import Soil Characterization Data (Later Phase)*

Importing supporting soil characterization data require geographic location or identifier. Soil characterization data can update or replace existing data. Replace existing data when the new data represents corrections in data results. Update and version the existing data when the new data represents a different time of observation of lab results at the study location. **Note: What are the geographic coordinates – Lat/long coordinates?**

2.4.5 Site-Pedon and characterization data joined (Later Phase)

In staging server authorized user joins Site-Pedon with associated soil characterization data. In cases where the data from LIMS have the original sampled as Site-Pedon data and data from NASIS represent the current up-to-date description both Site-Pedons are linked to the characterization data and appropriately referenced (sampled as, current).

2.4.6 Soil spatial data and tabular data are joined.

Authorized user joins spatial and tabular data. (Potentially a more automated process where User State Soil Scientist selects soil survey area spatial data and tabular data, system processes linkage).

- Spatial and tabular are matched on a specific area symbol.
- The map symbol in the spatial data is linked to the map symbol in the tabular data. Null musym are rejected.
- The mukey in the spatial data is added or revised using the mukey in the mapunit table.
- A successful join is when all polygons with a valid musym have a link to the tabular data with mukey. Internal soil survey polygons that are outside the survey have a musym BLANK and do not need a mukey (outside limit of survey). **Note: More work need on polygons labeled with non-mapunit symbols, such as DA (Denied Access), Dam, Levee, W (Water), M-W (Miscellaneous Water). It is very conceivable that DA could also be a valid mapunit symbol. NSSH may need updating.**

Unsuccessful Join

- User is notified of join problems.
- User attempts join with different selected spatial or tabular data.
- User returns to NASIS or Digitized Soil data and edits data.
- User re-exports from transactional databases.

2.4.7 Import area boundary maps

Classes of area boundary maps are used as layers to identify different types of soil data, their availability and their geographic extent. Initially the primary class is the official area boundary map based on the non-MLRA soil survey area boundaries and SSURGO status map. Publishing official soil survey data is not limited to non-MLRA area type, however each official soil survey boundary must be contiguous with adjoining survey areas and with no overlaps. This provides a single layer coverage of official soil survey data. For any given geographic location a customer can easily identify the current official soil survey data. Special use soil surveys are attributed on a different area boundary map. Overlapping areas are okay for special use soil surveys. Each special use soil survey area will be attributed with data clearly explaining the purpose of this special use soil survey, it's

limitations and constraints. Digitized area boundary maps that represent the geographic boundaries of official soil survey data and special use soil survey data are imported into the staging server.

- State Soil Scientist in coordination with National Soil Survey Center and National Cartographic Geospatial Center (NCGC) manages the boundary map.
- Edits to boundaries are updated in digital form.
- The spatial soil survey area boundary polygons are transmitted to the staging server.
- Authorized user validates boundaries.
- Staging server area boundary map is compared to the current warehouse version. User reviews and approves the updates. The existing boundary map in the warehouse is versioned and new the boundary map is published to the warehouse.
- **NOTE: Need additional information on business process, business units, and authorized users for updating, managing, and approving boundaries.**

Import Boundary Map Fails

- Unable to import boundary map.
- User that manages the map is notified.
- User makes corrections and transmits map again.

2.4.8 Staging Server Data Administration

Authorized users who have submitted a dataset have the ability to delete that dataset on the staging server. This might be used when a user process a dataset other than they intended. It could also be used when a user process either the spatial or tabular portion anticipating that the other matching piece would shortly be imported to the staging server. If however, the other matching piece, spatial or tabular, will be months before being ready, the user might choose to delete their dataset until a later date. Data administration functionality is dependent upon system constraints of storage and performance.

2.5 Evaluate and Approve Quality of Data and Release to Warehouse

2.5.1 Evaluate tabular soil survey data

An authorized user can evaluate the completeness and accuracy of the soil survey data by utilizing NASIS warehouse interpretations and reports or the Microsoft Access Template reports. If the data is acceptable user approves data for release to the warehouse.

- Authorized user selects soil survey data and displays data in various reports to review and validate.
- User runs reports to confirm data quality and compare data with existing data in warehouse.

Data not complete or accurate

- User rejects data.
- Data are deleted from staging server.
- User returns to NASIS and edits/updates data.
- User returns to transactional database to export data again.

2.5.2 Evaluate spatial and tabular soil survey data

User evaluates the soil survey data using GIS tools to view soil data and interpretations spatially.

- Authorized user accesses soil survey data on staging server.
- User selects the soil survey area in staging server and adjoining soil surveys from warehouse database or a data mart.
- The survey data is processed to either a format for Web Soil Data Viewer Data Mart or the data is extracted for download for client version of SDV. Other methods of spatially viewing the data on staging server may be possible.
- Extracted data from staging server are tagged as not official data for review and testing only.
- User evaluates interpretation patterns and reasonableness for the survey area. User also evaluates join of interpretation and soil properties with adjoining soil surveys.

Data not complete or accurate

- User rejects data.
- The rejected data (tabular, spatial, or both) are deleted from staging server.
- User returns to NASIS or Digitizing Unit Database and edits/updates data.
- User returns to transactional database to export data again.

2.5.3 Verify area boundary maps

Official area boundaries are the responsibility of the State Soil Scientist in cooperation with NSSC and NCGC. The typical case is a non-MLRA soil survey whose boundaries are already included in the official status map. No further action is needed, whether the survey is being updated or entered for the first time. However, the survey in the staging server could be a new area, possibly a locally defined area that could overlap an older survey area boundary. If the new data are to be the official data within their boundary, older boundaries must be adjusted by the State Soil Scientist to remove any overlaps or gaps. Some soil map unit polygons in older areas will also have to be archived so that only the new ones remain in the current official soil map.

- Display area boundary from staging server over the official status map to check for overlaps or gaps.
- State Soil Scientist is responsibility for defining and maintaining official area boundaries in cooperation with NSSC and NCGC.

NOTE: More detail is needed, who has authority to make

these kinds of changes, what's the process, digitizing updates, etc?

- Modifications to older survey maps are brought into the staging server as a new version; they are not edited directly in the warehouse.
- Existing survey data that are geographically partially updated, for example an existing survey area is split into two soil survey areas. One portion is updated with new data published. The remaining soil survey data that was once official for the entire area is now only official for a portion. Metadata and versioning updates to describe the original soil survey as being only official for a portion of the geographic area.

2.5.4 Evaluate NCSS Soil Characterization Data (Later Phase)

Authorized user evaluates the joining of Site-Pedon and characterization data. User approves data for updating or replacing existing warehouse data.

- NCSS data approved for public warehouse.
- Other private customer data approved for private warehouse.

2.5.5 Approve data for delivery to warehouse

This is the process of officially approving the release of the soil survey data to warehouse.

- State Soil Scientist approves data for release to warehouse.
- Data is approved as current official soil survey data for use in FOTG or as other special soil survey data not for official use. The data approved as current official soil survey data (detailed soil survey for use in FOTG) automatically becomes available in the FOTG/SSURGO Data Mart. Other data sets are made available in the warehouse as special use soil surveys not for official use.
- Date the official soil survey data is approved is the beginning date stamp this dataset is to be used as official data.
- Previous versions of the soil survey are stamped with the ending date. Current versions display date first official to present, previous version show beginning and ending date range when they were the official data.

Approval rejected

- User rejects data.
- Data is deleted from staging server.
- User returns to NASIS or Soil Spatial Data and edits data.
- User begins process of export over.

2.6 Warehouse Database (D5)

The warehouse database is a Geospatial database (spatial and tabular) providing a single source of current official soil survey data. The warehouse database will meet the following business and system requirements.

- Data will be protected from loss or damage, using a backup and recover system.
- The database model is designed to hold any type of area legend released by authorized data owners.
- The database serves as the source for the official national coverage of soil survey data with the capability to query, report and interpret the soil data for any geographic area. To be implemented in a later phase.
- Metadata including the contents of the tabular data dictionary, spatial metadata and other information needed to meet metadata requirements such as FGDC.
- There is only one set of data that represents the current official data for any geographic area, regardless of the legend – area type.
- Data for an existing survey area in the warehouse will not be updated by the inclusion of data for other survey areas.
- Provide versioning capability for soil digital- and tabular data when an area has new data superseding previous data.

2.6.1 Importing data

After data have been reviewed in the staging server and approved, they can be moved into the warehouse. Linkages specified in the data model are created by automated procedures.

- Overlaps of the new map and existing warehouse map are reconciled as which data is used as official in the area of the overlap. The State Soil Scientist determines which dataset to use as official in the overlap area. Relationships are updated as necessary.
- The warehouse database is queried to determine if a previous version of any object is already checked in, and version numbers are assigned to each object in the import. (If an object with the same NASIS key and date of update is found in the warehouse, its warehouse internal key and version number are used for foreign key references; no new version is created. This situation will occur mainly with data from lookup tables such as plants, geomorphic features, areas, etc.).
- New unique internal keys are assigned as needed using the numbering conventions from the warehouse database, and foreign key references are updated.
- The new map unit polygons are added to the current map unit feature class and linking tables are updated as necessary.
- Other digitized features are added to the appropriate feature classes in a similar manner.
- Tabular data are moved to the warehouse database and linked in.

- The processes for exporting to the data marts are triggered.

Import fails

- Notify State Soil Scientist.
- Clean-up data, roll back transaction.

2.6.2 Version and archive data

For each official soil survey area there can be only one set of current data. When an updated dataset comes into the warehouse it must be identified as the current version, and older data must be identified as superceded.

- The effective date for each object is recorded in the database. If needed, the expiration date for an object could also be recorded.
- Linkage tables in the database are updated to associate the versions of data available for a given geographic location, or for a link to a lookup table.
- In the official soil map coverage, polygons that have been superceded are tagged as archive layer so there will be only one current official polygon for a location.

2.6.3 Amending and updating existing official soil survey data (Later Phase)

This capability is updating an existing set of official soil survey data when only additional interpretations are needed with no change to existing chemical or physical data or existing interpretations. (See section 7 Soil Reports and Interpretations Data Mart)

- Authorized user publishes interpretation criteria to warehouse (See 2.4.2).
- State Soil Scientist identifies existing official soil survey to add additional interpretations.
- State Soil Scientist initiates update.
- The new interpretations are generated against the official soil survey warehouse data.
- The resulting generated interpretations and the existing official data are published to the staging server along with spatial data for the survey.
- State Soil Scientist validates and reviews the updated data set in the staging server.
- Approved data is published to warehouse and updates soil survey versioning.
- Extract existing spatial data for refresh of the tabular data.

Interpretation generator fails

- User selects a different interpretation.
- User edits interpretation in NASIS and re-published corrected version to warehouse.
- User quits.

2.6.4 Import local Microsoft Access template

Locally maintained and managed Microsoft Access Templates for tabular data import are published to the warehouse and geographically referenced for applicability.

- User transmits their locally maintained Microsoft Access template to the warehouse.
- User identifies the geographic area to which their template is applicable.

2.7 Export Soil Data to FOTG and SSURGO Data Mart

Any update to an official soil survey dataset in the warehouse triggers a process to export the latest data to the data mart. SSURGO data sets or digital datasets are created where the soil survey data include soil spatial data, tabular data, and metadata. Digital spatial datasets are those soil surveys that have spatial and tabular data that are released as preliminary data. Examples are soil surveys in progress (initial or update) where data is periodically released to the field office and public. FOTG data sets may include SSURGO, digital spatial data, but also may be just tabular data when digital soil data are not available.

- Import of a soil survey into the warehouse identifies the set of data to be exported. A FOTG export always contains the complete set of map units, components, interpretations, etc. that are checked into the warehouse.
- If a spatial data is included with the survey, its features are selected out of the current dataset and converted to one or more file formats for inclusion in the SSURGO package.
- The tabular data are exported using a process similar to the SSURGO Export option of NASIS, which creates a series of files in a format that can be loaded into Microsoft Access or other databases. The data model of the export is defined by the SSURGO standard.
- All export files are compressed into a single package, plus documentation files, and stored in the FOTG/SSURGO Data Mart.
- Associated metadata linked to soil survey area are stored in FOTG/SSURGO Data Mart.
- Catalogs are updated to reflect the presence of a new survey area

3 FOTG and SSURGO Data Mart

The FOTG and SSURGO data mart is the source for customer access of soil survey data for use in the Field Office Technical Guide, Section II and the Soil Survey Geographic Database. Access is through a web site that provides the ability to choose an area of interest and the types of data needed, then either download the data or order a CD. The foundation for data storage is a geodatabase. However, each soil survey will be prepackaged into SSURGO or FOTG product to increase efficiency in data delivery for entire soil survey area.

- ❑ Access for current official soil survey data.
- ❑ Access to archived versions is not required in real-time.
- ❑ Access to special use soil survey data is available.
- ❑ Connections to technical references, and customer information to aid in the selection, how to use and download of information.

3.1 Catalog of Available Data

As data is posted into the warehouse data mart the soil survey availability is automatically cataloged for Resource Data Gateway and for attributing navigation maps and choice lists.

- Catalog lists the current official soil survey data and archived versions and the dates that they were current.
- Catalog lists other special soil survey data.
- New soil survey data updated since some date is a reporting function. (See section 7 Soil Reports and Interpretations Data Mart)

3.2 Select Data

Customer selects soil survey data by navigating to a geographic area of interest, submitting geographic area reference, or by explicitly identifying the soil survey area. Initially this data mart only provides access to “traditional” soil survey data and not to supporting data (site/pedon, lab, images, etc).

3.2.1 Geographic navigation

Navigation starts with a view of the lower forty eight states of the United States, with map links to Alaska, Hawaii, Pacific Basin, Puerto Rico and Virgin Islands.

- Initial navigation aids are state boundaries.
- As customer zooms in official area boundaries and data availability become visible. Data availability statuses are: not currently available, available soon (contact information), tabular data only, and spatial and tabular data.
- Customer can toggle other special soil survey area maps on to display other data available.
- When customer has navigated into area of interest they initiate order placement.
- Select data by geographic reference. Selection of area is made by submitting area boundary in geographic reference (bounding box

or rubber band box) of decimal degrees for North American Datum 1983. **NOTE: reference system is constrained to the North American Datum 1983 in units of decimal degrees. Exception, Hawaii and potentially Pacific Basin use a different Datum.** Search for spatial data intersected by shapefile. Select feature individually. Select features intersected and contained by rubber band box.

3.2.2 *Select data by soil survey area*

Selection is by either explicitly entering soil survey area identification or choosing state and survey area from choice lists.

- Customer is given option to view the geographic extent of the survey area selected.

3.2.3 *Select data by date*

Provide capability to select data that have been updated since some date.

3.3 *Order Data*

3.3.1 *Customer places order*

- Geographic navigation may show several soil surveys available for area of interest.
- Customer can select all or a specific soil survey area.
- Catalog explicitly identifies soil survey data as current official soil survey data, archived version, or special use soil survey not for official use.

3.3.2 *Customer selects SSURGO data format*

Customer selects data format. **NOTE: need to confirm these are the supported delivery formats.**

- ArcInfo Coverage.
- ArcInfo Interchange File (e00).
- ESRI ArcView shapefile.
- Tabular data is zipped ASCII, pipe delimited, double quote text fields with each record terminated with a carriage return/line feed.
- Metadata

3.3.3 *Customer selects Microsoft Access template*

Customer is given option to download Microsoft Access Template.

- Customer can select About Template link to get a short description of what the Microsoft Access Template is about and why they may choose to download.
- For soil survey areas with tabular data only or spatial and tabular data in ESRI ArcView shapefile format the customer is given the option to download the MS Access Template.

- For soil surveys not in areas where local templates are available customer is given option to download national template.
- For soil surveys for areas where a local template has been identified customer is given the option to download the local template (preferred option) or national template.
- Selecting to download template includes template instruction guide for loading tabular data and contact information for local template and national template.
- Data is automatically loaded in the selected template at time of processing.

3.3.4 Customer selects delivery mechanism

- Customer selects either internet download or delivery on CD-ROM through the US mail, FedEx, UPS, etc.
- Delivery by mail will require customer entering mailing information: name, address, city, state, country, zip code.

3.4 Register for Notification of Data Updates

Customer is given the option to be notified if the data they have downloaded has been updated.

3.4.1 The customer information to record are:

- First and last name (optional).
- Organization (optional).
- Email address (required).
- Customer is given the option to register their information so each time they enter the site to download data they do not have to re-enter the information.

3.4.2 Customer has the ability to delete notification

- Customer identifies either a specific soil survey area that they no longer want to be notified of updates or complete removal all their notification information.

3.5 Deliver data

Process data request, formatting options, and delivery mechanism.

- FTP download process when complete Emails the customer when their data order is ready and provide download location.
- Mail CD ROM.

4 Data Flow Diagrams

The following data flow diagrams show the basic business process and data flow for soil data delivery and distribution.

4.1 Legend

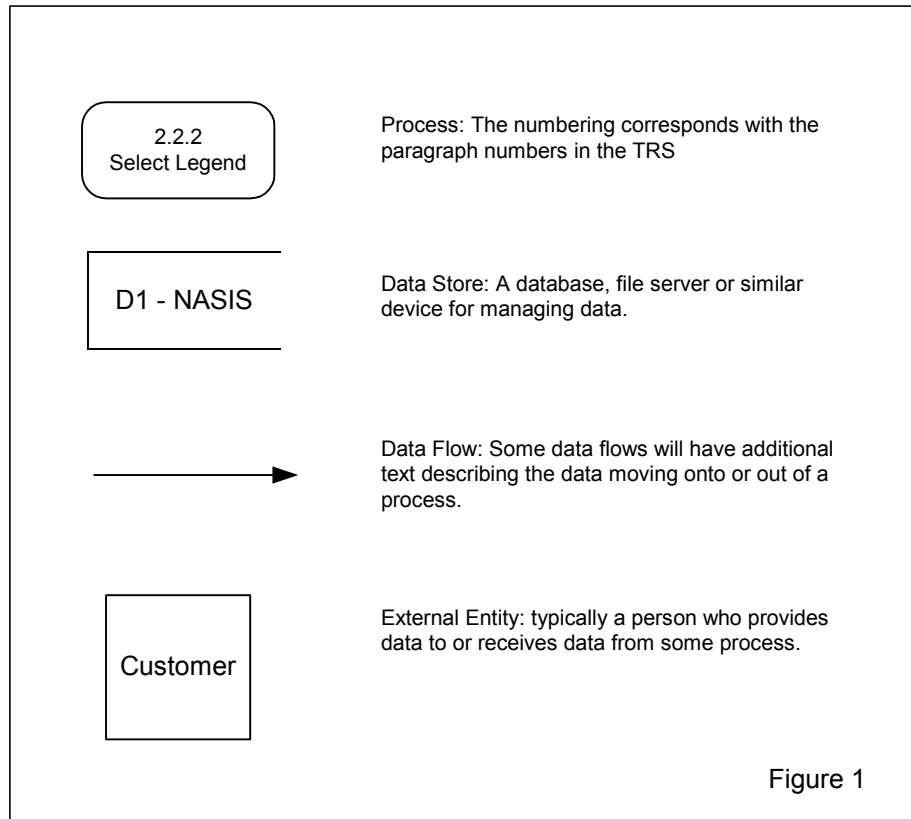


Figure 1: Legend of symbols used in Data Flow Diagrams (DFD)

4.2 Importing Soil Survey Data into the Warehouse

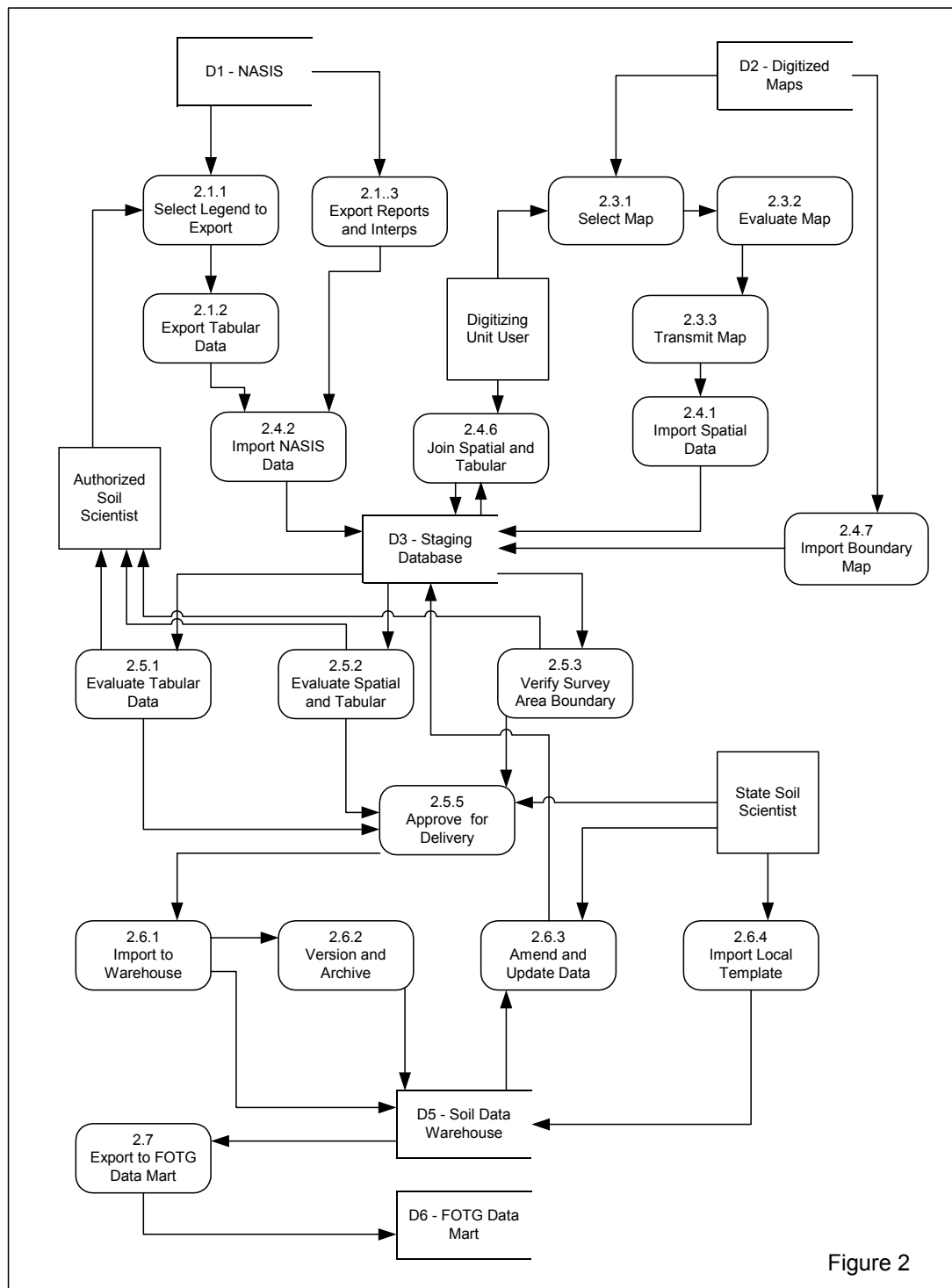


Figure 2

Figure 2: Data Flow for importing soil data into warehouse and delivering data to FOTG Data Mart.

4.3 Importing Laboratory Data into the Warehouse

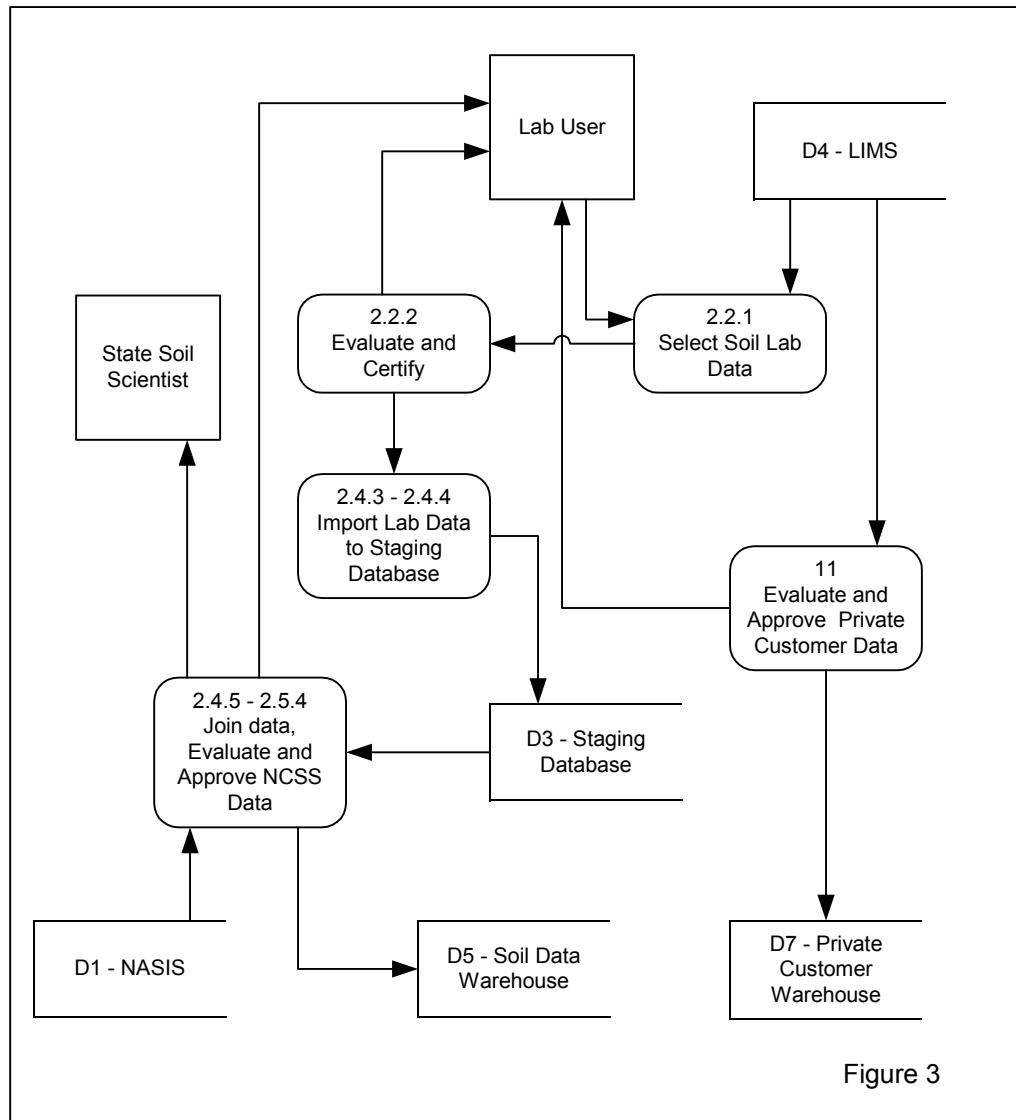


Figure 3: Soil Characterization Data Flow from Laboratory Information Management System (LIMS) to Warehouse.

4.4 FOTG and SSURGO Data Mart

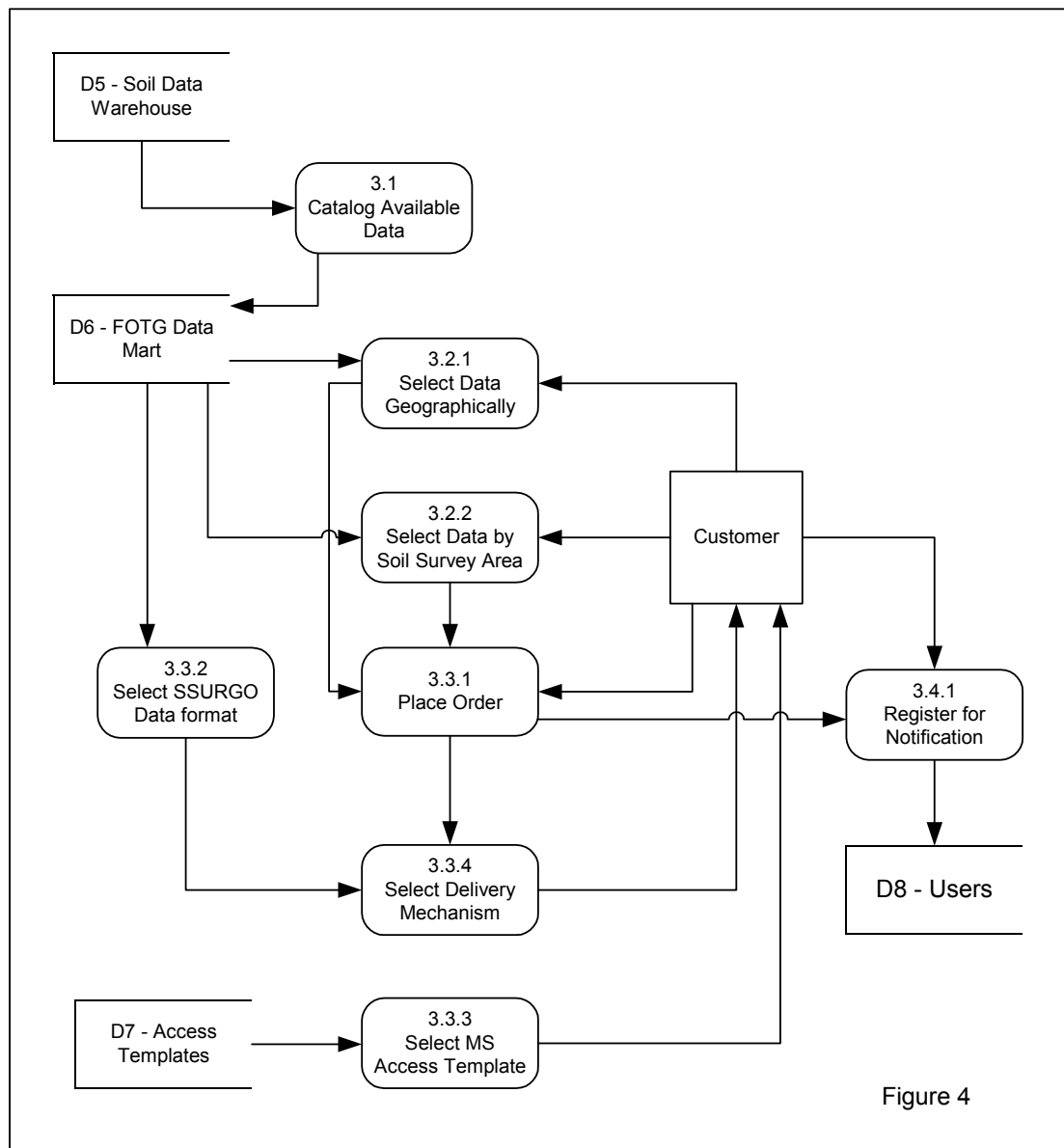


Figure 4: Data Flow FOTG/SSURGO Data Mart.

5 Data Model (Not Complete)

The Warehouse data model is based on the NASIS database in order to leverage NASIS code and capabilities of interpretation and report generation.

The main differences in the database model between NASIS and the warehouse are the elimination of management attributes (soil survey schedule) and the additions of the interpretation table (icomp), hydric rating and criteria, and versioning keys and attributes.

5.1 Business Rules

Relationships in the data model are designed to support the business rules listed.

5.1.1 Current and Archival Versions of Data

- Tabular data are archived in units consisting of a hierarchy of linked records. Individual records within the hierarchy cannot be updated in the warehouse without creating a new version of the whole archival unit. Archival units include:
- **Legend:** includes all map units along with their data map units, components, and interpretations, as selected in the NASIS export process. An update to interpretations, even without changing any other data, is considered a new version of the legend. For references to lookup tables (Plant, Geomorphic Feature, Ecological Site, etc.) the link is made to the version of the lookup record in effect at the time the legend version is checked in.
- **Area:** includes the area type, area and text records, although versioning takes place at the Area level. Certain area types, such as soil survey areas, have boundary polygons stored in the warehouse, so the area record and its boundary must be kept in synch.
- **Plant:** each plant record, with its synonyms and occurrence records, is archived independently. When a Plant is revised the new version does not need to be added to the warehouse until new versions of its associated local plants are checked in.
- **Geomorphic Feature, Ecological Site, and Other Vegetative Classification:** similar to Plant, each record is archived independently.
- **Pedon:** includes horizons and all child tables. When a Component is checked into the warehouse that has a link to a pedon, the current version of the pedon must be in the warehouse or the component-pedon link will be erased (a warning of this condition will be generated in the staging server).
- **Site:** Need to determine if changes to a site should be archived, and if a site observation can be added without creating a new version of the site. If sites and observations are versioned then a more complex relationship to Pedon is created. Each version of a pedon can link to one or more versions of a site and observation,

and each version of a site or observation can link to zero or more versions of a pedon.

- **Rules:** when a new rule or a modification is checked in, all associated evaluations and properties will be brought up to the latest version. The rule will retain links only to the evaluations and properties in effect at the time it was checked in.
- **NASIS User and Group:** not included in the warehouse. Columns in NASIS tables that reference user and group are not exported to the warehouse.
- **Documents and Images:** include Word, PDF, or other document formats, photographs, block diagrams, and other images. Can be stored as a non-text object or as a reference (URL) to an external file. They can be archived independently and are linked to specific records in the tabular database, such as a legend, map unit, site or pedon. These are many-to-many relationships.
- **LIMS Data:** need to determine the archiving requirements. Tentatively, it appears that Person and Organization might not need to be versioned; Site and Pedon should be merged with the corresponding NASIS tables and handled the same way; Project and Layer could be included in one archival unit independent of Pedon; Procedures and Analytes should be versioned; Results are not versioned but are linked to specific versions of Layer, Pedon, Procedure and Analyte.
- Spatial layers, such as the official soil survey boundaries or official soil map units, will be organized so that for any given time it is possible to determine unambiguously a single polygon that includes a given point on the ground. When a dataset is revised, the new polygons are incorporated into the current official layer and prior polygons covering the same locations are flagged as obsolete. State Soil Scientist is responsible for defining and maintaining soil survey area boundaries in cooperation with NSSC and NCGC. **Note: need more clarity?**
- The tabular data for a legend may be revised without changing the polygons in soil map layers such as SSURGO and STATSGO. The relationship between the spatial and tabular data must be maintained in such a way that a match can be made between a set of polygons and the specific map unit records applicable at a given point in time.

5.1.2 Identifying Official Data

There are several spatial layers that represent current official soil survey data, the soil survey area boundary layer, soil delineation (polygon, point, linear) layer, point special feature layer, and linear special feature layer. The point and linear special features are the spot symbols used on soil survey maps. The tabular data linked to these spatial features must be consistent with each other. Point and linear soil delineations are treated the same as polygon soil delineations. In addition it must be possible to identify the current official data using only tabular data, and

determine where data are missing even where there is no digitized soil data. Business rules such as the following will be enforced in the warehouse to accomplish this. **Fortner – Some will argue that STATSGO is also official, just a different scale.**

- Official survey area boundary polygons should all be linked to Area records of the same type, such as “Non-MLRA Soil Survey Area”. There should be a one-to-one match between the current versions of this area type and survey area polygons.
- Each current Area record for an official survey area should be linked to no more than one current Legend. If a new legend for an official area is entered in the warehouse it becomes the official legend and the old one becomes an archive version, regardless of the legend status. This should be verified by the user (State Soil Scientist) in the staging server.
- The map unit polygons within the survey area boundary must be linked to the map units in the current official legend. In the staging server tests will be run to verify that all map units and polygons are accounted for.

5.2 Tables and Attributes (Not Complete)

Generally the tables and attributes between NASIS transactional database and the warehouse database are the same with a few exceptions.

5.2.1 Business Answers

- Support of FOCS is NOT needed in the warehouse.
- Legacy interpretations are NOT published to the warehouse.
- Obsolete attributes (ordination symbol, component wildlife and forestry attributes, etc.) are published to the warehouse.
- Acreage data and overlaps are published to the warehouse.

5.2.2 New Tables and Attributes in Warehouse

- New generated interpretation result table (icomp).
- Hydric rating and criteria (reason) attributes.
- Data versioning attributes.

5.2.3 NASIS Attributes Not in Warehouse

The main attributes not included in the warehouse deal with program management data for the soil survey schedule and NASIS security (group, user), correlation decisions. **Note: Soil Survey Division still working on final decision on which attributes are not included in the warehouse.**

6 Web Soil Data Viewer Data Mart (Not Complete)

This data mart provides interactive access to the official soil survey data to select soil survey data geographically and create soil thematic maps that can be printed or saved. Soil survey data available for user viewing and processing are the State Soil Geographic (STATSGO) database and current official soil survey data as the Soil Survey Geographic (SSURGO) database and tabular only data as part of Field Office Technical Guide (in SSURGO tabular format). The functionalities are:

- ❑ Navigation facility, with aids in navigation becoming active at appropriate scale for such as major highways, county roads, hydrography, orthophotography.
- ❑ Capability to zoom, pan, identify, set home extent, reset to beginning.
- ❑ Data availability is indicated as: not currently available, available soon (contact information), tabular data only, and spatial and tabular data.
- ❑ Processing capability for wider area analysis. What is the constraint? Certainly processing data for all the US may not be appropriate with SSURGO data, but what about for a State or Region. Would a state prime farmland or hydric map be appropriate? What are the system limits?
- ❑ Ability to add additional themes such as hydrography, elevation, climate, etc. for multiple theme analysis. What are system constraints for processing multiple layers?
- ❑ Ability to attribute polygons with labels.
- ❑ Ability to print and save map.
- ❑ WSDV ability to process local and national interpretations. Is it interactive showing only those choices that are available in the data?
- ❑ Ability to process several themes with WSDV.
- ❑ Ability to extract and download extent and themes created.
- ❑ Ability to report tabularly. Possible linkage to reporting capabilities for selected features with reports from reporting data mart. At a minimum reports generate for processed thematic map, with attribute definitions and processing methods used.

6.1 *Navigate*

Customer selects soil survey data by navigating to a geographic area of interest or by explicitly identifying an area of interest by location place name (city, county, state, soil survey area, PLSS, etc.).

6.1.1 *Geographic navigation*

Navigation starts with a view of the lower forty eight states of the United States, with map links to Alaska, Hawaii, Pacific Basin, Puerto Rico and Virgin Islands.

- Initial navigation aids are state boundaries.
- As customer zooms in official survey area boundaries and data availability become visible. Data availability statuses are: not

currently available, available soon (contact information), tabular data only, and spatial and tabular data.

6.2 *Show data availability scale dependency (STATSGO/SSURGO)*

As customer zooms –in or zooms-out the navigation view displays geographic area references of data availability. At a national scale STATSGO is shown available and a closer regional or State scale both STATSGO and available SSURGO data are displayed. At a soil survey area or county scale SSURGO data is displayed as available.

6.3 *Web Soil Data Viewer Interface*

The Soil Data Viewer interface displays a list of available interpretations, processing options for creating a map or tabular report. STATSGO and SSURGO data availability is scale dependent.

- List of available interpretations are dependent on the scale and geographic area selected. At a small map scale only interpretations for STATSGO are available. As the user zooms in to a larger map scale the interpretations for each soil survey (SSURGO) in the selected area are available. Interpretations that are not in the soil survey data are not displayed as a choice.

7 Soil Reports and Interpretations Data Mart (Not Complete)

Analysis will be completed in Total Requirements Statement Phase 2 at a later time.

The soil report and interpretation data mart provides the capability to run a report or interpretation for a selected soil survey area, supporting pedon or characterization data (Lab data). Select capability by geographic area, complete or portions of soil survey areas, or by map unit or component attributes.

- ❑ Authorized user imports and manages ready to use reports.
- ❑ Customer selects from set of available reports.
- ❑ Customer navigates to area of interest.
- ❑ Ability to run report on selected area of data and selected map units.
- ❑ Ability to run tabular reports with soil data (WSDV?).
- ❑ Customer-friendly interface that allows customer to run a report for desired interpretations without needing to know which report name contains the desired interpretation or property. This might be a description of the report describing its major contents, or an optional reporting system interface that leads the customer to select the report for their needs.
- ❑ Ability for authorized user (State Soil Scientist) to update an official set of data with additional interpretations. The selected interpretations are run against the current official data and processed to staging server. User reviews and approves data. The survey data is republished as an update and new version of the survey area. This provides State Soil Scientist with the ability to add additional interpretations without the risk of updating existing interpretations or physical and chemical properties already approved.
- ❑ Query capability to select legends and map units by selected attributes for reporting or interpretation purposes.

7.1 Select Data

Customer selects soil survey data by navigating to a geographic area of interest or by explicitly identifying the soil survey area.

7.1.1 Geographic navigation

Navigation starts with a view of the lower forty eight states of the United States, with map links to Alaska, Hawaii, Pacific Basin, Puerto Rico and Virgin Islands.

- Initial navigation aids are state boundaries.
- As customer zooms in official survey area boundaries and data availability become visible. Data availability statuses are: not currently available, available soon (contact information), tabular data only, and spatial and tabular data.

- Customer can toggle other special soil survey area maps on to display other data available.
- When customer has navigated into area of interest they select report or interpretation.

7.1.2 *Select data by soil survey area*

Selection is by either explicitly entering soil survey area identification or choosing state and survey area from choice lists.

- Customer is given option to view the geographic extent of the survey area selected.

7.1.3 *Select specific map units for soil survey data*

Provide the customer the ability to specify all or select specific map units for reporting or running interpretations.

7.1.4 *Select report or interpretation*

Provide customer the ability to select from the available reports and interpretations to run against selected soil survey data and view, print and save output of report or interpretation. Customer can select report or interpretations by:

- Report or interpretation name.
- Report or interpretation subject content.

7.1.5 *Comparing Data*

Provide the ability for customers to compare differences in the current official soil survey data and previous versions of data. Provide the ability to identify geographic areas where soil survey data have been updated since some date. Provide authorized user the ability to compare differences in current official soil survey data and most current data in the managed transactional database (NASIS).

Comparing two versions of data

- Customer selects the primary survey data of interest.
- Customer selects the matching survey data of previous versions to compare.
- Customer initiates compare report.
- Report results are either viewed on screen or emailed to customer.

Comparing new data available since some date

- Customer selects geographic area of interest, soil survey area, state, region, etc.
- Customer inputs date range, starting to ending date to find soil surveys that have new data between starting and ending date.
- Report results are viewed spatially or as tabular report showing the soil survey areas, the version dates, and when updated date.

7.1.6 *Updating Official Soil Survey Data*

Capability for authorized user to update an official soil survey area in the warehouse with the addition of new interpretations without risking the updating of existing interpretations, or physical and chemical properties. This process is similar to the initial process of publishing a soil survey from NASIS, except that the processing of the new interpretations is from the current existing soil survey data in the warehouse. Authorized user is State Soil Scientist.

- Select soil survey area.
- Select interpretations to add to the current official soil survey data.
- Initiate interpretation generation.
- Soil survey data with new additional interpretations are processed and imported into staging server.
- Review and validate the updated soil survey data.
- Approve update soil survey data for release to warehouse.
- Approved data moves to warehouse database creating new updated version of the current official soil survey data. Existing soil survey area data are versioned and no longer the current official data.

7.2 *Register for Notification of Data Updates*

Customer is given the option to be notified if the data they generated reports or interpretations have been updated.

7.2.1 *The customer information to record are:*

- First and last name (optional).
- Organization (optional).
- Email address (required).
- Registering of customer information gives the customer the option of recording their customer information so each time they enter the site to download data they do not have to re-enter the information.

7.2.2 *Customer has the ability to delete notification*

- Customer identifies either a specific soil survey area that they no longer want to be notified of updates or complete removal of all their notification information.

8 Custom Export Data Mart (Not Complete)

Analysis will be completed in Total Requirements Statement Phase 2 at a later time.

The custom export data mart that provides the ability to select only those soil characteristics desired for a limited list of map units. The initial selection is by submitting a geographic area or identifying a specific soil survey area. The user then has the capability to select the desired map units and/or soil attributes for the area.

Referential linkage of selected attributes is enforced.

- ❑ Navigation capability to find area of interest.
- ❑ Query and select capability.
- ❑ Identify desired tables and attributes. Force table relationships so a customer can't select to download crop yield without getting component, map unit and legend identifiers.
- ❑ Ability to select download format, ASCII text, quote text, field delimiter, etc.
- ❑ Customer registers notification information when process complete, size packaged data and download or shipping information.
- ❑ Data is packaged for FTP download or CD-ROM shipment.
- ❑ Data is packaged with soil survey versioning information, metadata processing information, etc.
- ❑ Initially soil survey data is selected by entire or complete soil survey area. A future enhancement providing clipping/cutting specific geographic areas.

Custom access is through a web site that provides the ability to submit or choose an area of interest and the types of data needed, then either download the data or order a CD-ROM.

- ❑ Access to official soil survey data.
- ❑ Access to archived versions is not required in real-time.
- ❑ Access to special use soil survey data is available.

8.1 Web Access Interface

The web access interface needs to provide the capability to find available soil surveys for a specific geographic area, with connections to technical references, and customer information to aid in the selection and download of information.

8.2 Catalog of Available Data

As data is posted into the data mart the soil survey availability is automatically cataloged for the Resource Data Gateway and for attributing navigation maps and choice lists.

- Catalog lists the current official soil survey data and archived versions and the dates that they were officially.
- Catalog list other special soil survey data.

8.3 Select Data

Customer selects soil survey data by navigating to a geographic area of interest, submitting geographic area reference or by explicitly identifying the soil survey area.

8.3.1 Geographic navigation

Navigation starts with a view of the lower forty eight states of the United States, with map links to Alaska, Hawaii, Pacific Basin, Puerto Rico and Virgin Islands.

- Initial navigation aids are state boundaries.
- As customer zooms in official survey area boundaries and data availability become visible. Data availability status are: not currently available, available soon (contact information) tabular data only, and spatial and tabular data.
- Customer can toggle other special soil survey area map on to display other data available.
- When customer has navigated into area of interest they initiate order placement.

8.3.2 Select data by soil survey area

Selection is by either explicitly entering soil survey area identification or choosing state and survey area from choice lists.

- Customer is given option to view the geographic extent of the survey area selected.

8.3.3 Select data by geographic reference

Selection of area is made by submitting area boundary in geographic reference (bounding box or rubber band box) of decimal degrees for North American Datum 1983. **NOTE: reference system is constrained to the North American Datum 1983 in units of decimal degrees. Exception, Hawaii and potentially Pacific Basin use different Datum.**

- Search for spatial data intersected by shapefile.
- Select feature individually.
- Select features intersected and contained by rubber band box.

8.3.4 Select data by date

Provide capability to select data that have been updated since some date.

8.3.5 Select data attributes

Customer selects which tables and attributes of interest for the selected soil data. Table relationships are enforced, such that if component horizon attributes are selected the minimum component, mapunit and legend data to uniquely identify and reference the horizon data are automatically selected.

8.4 Order Data

8.4.1 Customer places order

- Geographic navigation may show several soil survey datasets available for area of interest.
- Customer can select all or specific soil survey data.
- Catalog explicitly identifies soil survey data as current official soil survey data, archived version, or special use soil survey not for official use.

8.4.2 Customer selects delivery data format

Spatial data

Initially the spatial data formats are pre-packaged. Future enhancements are ability to format clipped/cut geographic areas into delivery format.

- Modified Digital Line Graph (DLG-3) in NAD 27 or 83.
- ArcInfo Coverage.
- ArcInfo Interchange File (e00).
- ESRI ArcView shapefile.

Tabular data

- Tabular data is zipped ASCII, pipe delimited, double quote text fields with each record terminated with a carriage return/line feed.

8.4.3 Metadata for custom data

Provide minimum metadata describing the spatial data, table and attributes selected, soil survey versioning information, and use constraints and limitations for partial data extract.

8.4.4 Customer selects delivery mechanism

- Customer selects either the internet download or delivery on CD ROM through the US mail.
- Delivery by mail will require the customer to enter mailing information: name, address, city, state, country, zip code.

8.5 Register for Notification of Data Updates

Customer is given the option to be notified if the data they have downloaded has been updated.

8.5.1 The customer information to record are:

- First and last name (optional).
- Organization (optional).
- Email address (required).

- Customer is given the option to register their information so each time they enter the site to download data they do not have to re-enter the information.

8.5.2 Customer has the ability to delete notification

- Customer identifies either specific soil survey areas that they no longer want to be notified of updates or complete removal of all their notification information.

8.6 Deliver data

Process data request, formatting options, and delivery mechanism.

- When FTP download process is complete, the customer is Emailed when their data order is ready and the download location.
- Mail CD ROM.

9 Application Programming Interface (API) Data Mart (Not Complete)

Analysis will be completed in Total Requirements Statement Phase 2 at a later time.

The application programming interface is a data mart of software interfaces for retrieving data from the warehouse database. This data mart provides the capability for interactive models or third party software packages to access and retrieve specific soil survey data for use in their application. This provides for real-time access to the current official soil survey data without the need for replicating the data in their application resulting in the potential of their data becoming out-of-sync with the official data.

- ❑ User guides and technical references that can be printed or download.
- ❑ Detailed step-by-step instructions on selection, processing, data transfer, etc.
- ❑ Soil survey versioning information and metadata of processed data included.

10 Integration with Other Resource Databases (Not Complete)

Analysis will be completed in Total Requirements Statement Phase 2 at a later time.

Integration with other resource databases can be integration at the interface level, such as the capability to navigate to a geographic area and view soil theme layer, climate layer, vegetation layer, etc.

Appendix A

Fundamental Requirements of Soil Data Delivery and Distribution

Appendix A

Fundamental Requirements

Excerpts from Soil Data Delivery and Distribution Draft Requirements Statement, October 2001.

The fundamental requirements for delivering soil data and information are organized and described by the requirement. Examples of the requirement are included in the description.

General

The basic requirement is to deliver consistent official soil survey data of high quality that meets our customer needs and national program responsibility. Soil data referenced here is most commonly the map unit information for a soil survey. However, supporting soil data such as point site pedon data and soil characterization data are included.

1 *Identify and Access Current Official Data for a Specific Use (implemented this phase)*

A wide variety of users, both internal and external, need access to the current set of soil survey data for land use planning and natural resource assessment. We want to assure that the data are current, consistent, and repeatable. We also want to ensure that the data they acquire are the data we have reviewed and approved for general release. For example, district conservationists in field service centers need access to officially recognized soil survey data for conservation planning and administration of conservation programs. This data must be the current official data. We have typically satisfied this need by maintaining current official data in the FOTG, but the FOTG is not generally available to all users of soil survey data. We need the ability for all users, internal and external, to identify and access current official data whether they get this data via the Web, from a State Soil Scientist, or from a local field office. Each distribution point should be delivering the same data.

- ☐ Current official data for modern conservation planning.
- ☐ Official data for national program implementation.

2 *Identify and Access Previous Versions of Official Data for a Specific Use (implemented this phase)*

In the administration of agency conservation programs, and particularly for NRI, we need access to previous versions of official data. We also need to know the time period during which the data were current. For example, NRI conducts periodic inventories of natural resource conditions and uses soil survey data to enhance the value of the inventory. These periodic inventories are analyzed to determine trends in resource conditions. To perform accurate trend analysis, NRI needs access to versions of data appropriate to each inventory. Likewise, the administration of CRP and other agency conservation programs require reference to official data current at the time contracts are executed.

- ❑ Some data can be restricted to authorized access. This might occur where soil data has been published with significant errors. The data would be archived but unavailable for general access by users.
- ❑ Address the linkage to different versions of data, component in a DMU linked to a pedon, the original sampled as pedon and if updated the new pedon. The DMU needs to link the pedon data the user needs for a specific use.
- ❑ Example is the archived data (frozen 1990 list) for programmatic requirements (CRP). Implementation in electronic format is not required but it's continued access availability is still required (e.g. hard copy in FOTG).

3 *Maintain, Identify and Access More Than One Set of Official Data for a Geographic Area (implemented this phase)*

Typically, we maintain two sets of official data for any geographic area; detailed soil survey data, and general soil map data. In many instances, these are SSURGO and STATSGO data. Both are official data, and both are good products when used as intended. Likewise, many states have developed statewide legends designed to meet specific needs such as equalized taxation or equitable land use regulation. The legends and associated data used by these states are designed to meet specific needs, which may be different from the needs identified for individual survey areas. Thus, we need to be able to:

- ❑ Maintain, identify, and access one or more soil survey data sets appropriate for different specific uses in a given geographic area.
- ❑ Uniquely identify legends, map units and components (NRI).
- ❑ Frozen soil list e.g. 1990 frozen HEL list (CRP).

4 *Maintain, Identify and Provide Access to Soil Survey Supporting Data for a Geographic Area (implement in later phase)*

Supporting soil survey data include site pedon descriptions collected during conduct of soil survey and technical soil services, soil characterization data, soil performance measurement data, photographic and graphic images, and field observation notes.

- ❑ For a geographic area identify supporting soil data and provide access to utilize these data in reports, interpretations, and downloads with other resource data.

5 *Access to Reporting and Downloading Capability (implement in later phase)*

Majority of the other requirements deal with the identification and access to soil data. Access implies that the user has the ability to get soils data and information as a tabular report, a spatial map, or a prepared product like the published soil survey report. User can view or get as an electronic version.

- ❑ Soil maps and soil thematic maps which include the ability to select and create a specific interpretative map. Examples are Septic Tank Absorption Field or weighted average organic matter content for the upper 10 inches of the soil.

- ❑ Tabular reporting includes viewing and downloading results from a selection of available reports. Reporting capability allows for selecting needed data.
- ❑ Support for new reports to meet user needs.
- ❑ Type of Reports include print on demand interim soil survey reports, published soil survey reports, map unit descriptions and brief non-tech soil descriptions.
- ❑ Download selected soil attributes in electronic format.
- ❑ Graphical Report such as properties with depth (clay/depth) and water table depth per month.
- ❑ Images (profile, landscape, land use photographs).

6 *Identify Changes Between Versions of Data Released to Users (implement in later phase)*

When data users discover differences in interpretive results between one version of data and another version, we need to be able to examine both versions of data and determine the differences that exist between the versions.

- ❑ If portions of a soil survey have been updated by updated map unit DMU, then we need to be able to identify which map units have updated data and report both tabularly and spatially.
- ❑ Identify changes in criteria for updated interpretations.
- ❑ Identify new interpretations included in the updated soil survey data.
- ❑ Identify which attributes have changed and the before and after values.

7 *Identify Geographic Areas Where New Data Have Been Updated Since Some Previous Date (implement in later phase)*

NRI needs to evaluate workload for sampling by examining where soil survey data have been updated since the last NRI cycle. Geographically select PSUs for areas that have updated soil survey data. Another example is review of the status of updated soil survey data in the nation.

- ❑ Report tabularly or spatially.

8 *Identify Changes Between the Most Up-To-Date Data and Official Data (implement in later phase)*

After soil survey data are examined, certified, and released for use as official data, we continue to “maintain” the data by correcting errors, supplement the existing data with new data, or in the case of updating a survey area make substantial changes to the data. State soil scientists are responsible for maintaining a current set of official data. They need to know what changes have been made in the NASIS transactional database (most up-to-date data) since the last official data were released. This provides a tool to assist in determining if a new version of official data must be released.

- ❑ Access to most up-to-date data is restricted to authorized users.
- ❑ A variation is the need to distinguish between components and map units with updated data (taxonomy) based on field investigations from those map units that have updated component taxonomy because a series has been updated but no field investigations have occurred on the map unit.

9 *Deliver the Most Up-To-Date Data for Specific Uses (implemented this phase)*

Most up-to-date data is in NASIS transactional database. Although most users need a known version of static official data with an identifiable citation, some users need access to the most up-to-date data not yet publicly released. This data includes map unit information, supporting soil survey data, interpretations, technical references, etc. Access to pre-released data meets a very specific need such as field-testing, model validation (WEPS) and field validating interpretation criteria and results. Another example is utilizing existing non-released data for a special site investigation such as soil properties on research plots and soil characteristics for engineering practices where users need the most up-to-date data for their specific use. The most up-to-date data normally is not used for official purposes.

- ❑ Access to most up-to-date data is restricted to authorized users.

10 *Deliver Data to Meet Specific Needs (partially implemented this phase)*

Different products and different data users have different needs for data. We need to provide soil survey data that have both format and content that meet the needs of these products and data users. SSURGO is a data product that has specific data needs that are different from the needs of erosion prediction or other environmental models. NRI may require that key fields and specific identifying characteristics (map unit symbol, component name, slope) be available to create choice lists for field inventory tools. Once the choices are entered in the field, NRI needs to re-visit the official data of the same version to acquire soil survey data used to enhance NRI analysis and products. Some applications, such as Customer Service Toolkit and NRI, need access to previous versions of data of known vintage (so-called “frozen data”). The following is a preliminary list of products.

- ❑ SSURGO.
- ❑ STATSGO.
- ❑ NRI data set.
- ❑ FOTG includes issues related to state custom MS Access template reports and Soil Data Viewer rules file.
- ❑ Published soil survey report.
- ❑ Soil characterization data.
- ❑ RUSLE2 data set.
- ❑ WEPS data set.

11 *Create New Interpretations from Current or Previous Official Data (implement in later phase)*

In cases where the physical, chemical, and morphological properties in the current official data have not changed, but new interpretations are needed, we want to be able to create new interpretive results from the existing current official data including soil survey supporting data. One example of this situation occurs when agency programs require a new specific interpretation, such as crop growth index, to be uniformly applied in all situations. The new interpretive result is not available for survey areas previously released as official data. We

do not want to make any changes in the soil properties in those data sets. In some cases, we may even want to create new interpretive results from previous versions of data in the administration of agency programs.

- ❑ Map unit level interpretation results based on dominant soil, dominant condition, most limiting, least limiting or weighted average.
- ❑ Ability to interpret using aggregated component or site point pedon data.
- ❑ Update official data with new additional interpretations.
- ❑ Interpretation utilizing other resource data such as climate, land use, vegetation, geology, political boundaries or cultural features.

12 *Apply Interpretive Criteria to Selected Map Units or Geographic Areas (implement in later phase)*

Some interpretive criteria are intended for use on limited geographic areas such as towns, counties, or watersheds. We want to select the specific map units, pedon sites or geographic areas on which to apply the interpretive criteria.

- ❑ Development of LESA for a jurisdictional area.
- ❑ Displaying an interpretation valid for a specific geographic area such as an interpretation based on county or state septic tank absorption field criteria.

13 *Provide Selected Attributes for Any Geographic Area (implement in later phase)*

When delivering data to meet specific purposes, we need to select only those soil characteristics desired for a limited list of map units. For example, when working with a consulting engineer on a residential development, we need to provide current official data for only the five map units on the tract under consideration and we need to provide soil moisture, drainage class, restrictive features, and construction interpretations for only those five map units.

- ❑ Source Technical Soil Service Analysis Document.

14 *Select Data by Any Attribute Without Respect to Geographic Area (implement in later phase)*

We want to select from the most current official data set those map units that meet certain defined characteristics regardless of their geographic location. One example of this function is to test the effect of new taxonomic criteria where the results are expected to be limited to a geographic region. The results of the selection may show that soils in an entirely different geographic region are also selected, indicating possible deficiencies in the new criteria. We may also want to select from the official data set those map units affected by new national conservation program rules.

- ❑ Provide spatial (map) and tabular reports and download capability.

15 *Provide Data That Can be Used to Create Seamless Spatial Coverage (implemented this phase)*

When official data are delivered for different soil survey areas, we want these data to join with their corresponding spatial data in such a way that data from the two survey areas can be used to create a seamless coverage. In many cases, an

area of interest straddles a survey area boundary and requires soil survey data from multiple soil survey areas. We want to join data from these areas easily and unambiguously to create a single seamless coverage that we can use for land use planning.

16 *Provide Complete National Coverage of Data (implemented this phase)*

Many nationally administered programs, such as the Conservation Reserve Program, require access to a national single coverage of current official data that is as geographically complete as possible. This includes the spatial and tabular (attributes) soil survey data. Typically, selected attributes are used in the administration of a variety of programs and in some cases, new interpretive criteria may need to be created or new results calculated from the existing current official data.

17 *Notify Data Users When Data Have Changed (implemented this phase)*

The National Soil Survey Handbook specifies that, to the extent possible, we notify data users when the data they have acquired has been updated. We also need to be able to identify the changes that have occurred (*see* Identify changes between versions of data released to users).

- ❑ Provide users the option of being notified when data they have accessed and download has been updated and changed. No requirement that user must leave any information when accessing and retrieving data.
- ❑ This covers the publicly released soil data not the most up-to-date transactional data.

18 *Notify Data Users of Product Plans and Progress (implement in later phase)*

When data users are told that their desired soil survey product is unavailable, they routinely inquire as to when it will be available or whether there are plans to produce the requested item. We want to inform data users of critical timeframes affecting the products they need.

- ❑ Provide access to soil survey status maps or other information that provides users with information about the progress, plans or product delivery schedules.

19 *Provide Stability in Product Content and Delivery Format (implemented this phase)*

Users of a specific product, such as SSURGO, rely on a stable structural content (tables, columns, relationships) and delivery format (ASCII pipe delimited text, XML, MS Access DB, etc) for a period of time. Provide a period of time (3 to 5 years) where a product will be available with the same structural content and delivery format. When a product structural content and delivery format are updated, provide users the ability to get soil data in the last version of the content and format. This allows users who have developed applications a period of time to re-tool their application.

- ❑ Provide stability in data transfer content and format. NRIS-Terra downloading soil survey data in SSURGO version 2 format can rely on access availability.
- ❑ Models using data extract routines continue to have access to the same data format.
- ❑ Provide users with a notice and time line for product content and delivery format changes.

20 *Eliminate Inconsistency (implemented this phase)*

We must guarantee that soil survey data used by one application is consistent with the same data used in another application or acquired from another source. For example, soil survey data used in the FOTG, Customer Service Toolkit and Soil Data Viewer must be consistent with data in SSURGO product. We expect that a user of SSURGO must see the same soil properties and interpretive results as a user of the Soil Data Viewer for the same vintage and version of the soil survey.

21 *Eliminate Redundancy (implemented this phase)*

To the extent possible, we must reduce multiple instances of what are intended to be the same data. Maintaining multiple instances of soil data that represent the official source makes it difficult to maintain consistency of the current official data.

22 *Protect Data from Loss or Modification (implemented this phase)*

Official data are static and must not be changed unless a new version is created. When new versions of data are made available, the previous versions must remain available and must be protected from loss. In a manner similar to the FOTG, once a data set is available and used in conservation planning, that version of the data must be available even after it has been updated by a newer version because contract provisions and management decisions have been based on the previous data.

23 *Comply with Federal Geographic Data Committee (FGDC) Soil Geographic Data Standard (implement in later phase)*

In the process of maintaining consistency and providing data to meet specific needs, at least one view of the data must comply with FGDC standards.

24 *Provide Access to the Technical References, Standards and Guides for Soil Survey (implemented this phase)*

Provide access to Soil Survey Technical References, Standards and Guidance for conducting soil survey. NCSS partners and general users use Official Series Descriptions and the Soil Classification File to understand the series concepts and soil landscape relationships. Technical references need to be available.

- ❑ Soil Survey Manual.
- ❑ Soil Taxonomy.
- ❑ National Soil Survey Handbook.
- ❑ Field Book for Describing and Sampling Soil provides guidance and conventions for describing soils.

- ❑ Soil Survey Laboratory Methods Manual.
- ❑ Forest Service – Terrestrial Ecological Unit Inventory Technical Guide.

25 *Provide Metadata, Detailed Information or Description of Products and Data Provided to Users (Partially implemented this phase)*

Metadata for any data products delivered to customers. Metadata includes the formal FGDC metadata standard for SSURGO, but also includes attribute data dictionary for selected data, description of products and their use constraints and limitations, methods used in processing and creating the product or data extracts, readme files, etc.

- ❑ Description of criteria used in generating interpretations.
- ❑ Criteria used in interpretative attributes (K, T, prime farmland, hydric, etc.) can be included in technical references, guides and handbooks. (see Provide Access to the Standards for Soil Survey).
- ❑ Versions of Metadata.
- ❑ Help on access, selecting and referencing official data.
- ❑ Information for points of contact for additional use and understanding of the soil survey information.
- ❑ Provide for user comments and feedback.

26 *Integration with Other Resource Databases (implement in later phase)*

Integration with other resource data sources provides the ability to create interpretations using soil data and other data sources. It also provides an improved level of data consistency and quality by reducing duplication of the soil data that is managed by other entities.

- ❑ Integrate with other resource data layers (climate, ecological site information (ESIS), vegetation, PLANTS, land use, political boundaries, cultural, etc.)
- ❑ Integrate with other soil data sources such as NRIS-Terra, ARS research plots, University soil data, and private sector.

Appendix B

Priority List of Fundamental Requirements

Appendix B

Priority List of Fundamental Requirements

Excerpts from Soil Data Delivery and Distribution Draft Requirements Statement, October 2001.

This table is designed to give guidance to the design and implementation in a phased approach to development. This list has been approved by Soil Survey Division.

Requirement	Business Area	Priority	Implemented This Phase
19 - Provide stability in product content and delivery format	USFS, SSURGO, Models	1	Yes
22 - Protect data from loss or modification	All	2	Yes
1 - Identify and access current official data for a specific use	FOTG, NRI, SSURGO	3	Yes
20 - Eliminate inconsistency	All	4	Yes
10 - Deliver data to meet specific needs	SSURGO, FOTG, Models	5	Partially
21 - Eliminate redundancy	All	6	Yes
2 - Identify and access previous versions of official data for a specific use	NRI, CRP	7	Yes
16 - Provide complete national coverage of data	National Program	8	Yes
5 - Access to Reporting and downloading Capability	All	9	No
11 - Create new interpretations from current or previous official data	National Programs	10	No
12 - Apply interpretive criteria to selected map units or geographic areas	LESA, National Programs	11	No
3 - Maintain, identify and access more than one set of official data for a geographic area	SSURGO STATSGO	12	Yes
15 - Provide data that can be used to create seamless spatial coverage	General	13	Yes
25 - Provide Metadata, detailed information or description of products and data provided to users	All	14	Partially

7 - Identify geographic areas where new data have been updated since some previous date	NRI	15	No
23 - Comply with Federal Geographic Data Committee (FGDC) Soil Geographic Data Standard	All	16	No
4 - Maintain, identify and provide access to soil survey supporting data for a geographic area	Tech Soil Services	17	No
13 - Provide selected attributes for any geographic area	Tech Soil Services	18	No
6 - Identify changes between versions of data released to users	General	19	No
17 - Notify data users when data have changed	General	20	Yes
9 - Deliver the most up-to-date data for specific uses	Tech Soil Services	21	Yes (NASIS)
26 - Integration with Other Resource Databases	All	22	No
24 - Provide access to the technical references, standards and guides for soil survey	All	23	Yes
8 - Identify changes between the most up-to-date data and official data	State Soil Scientist	24	No
18 - Notify Data Users of Product Plans and Progress	General	25	No
14 - Select data by any attribute without respect to geographic area	Tech Sers, National Programs	26	No

Appendix C

Design Unit Phases

Appendix C

Design Unit Descriptions

Excerpts from Soil Data Delivery and Distribution Outline Physical Design, December 2001.

The description of each design unit is in a summary form. Capacity estimates are preliminary, based on experience with existing systems such as SSURGO and MUIR. The next stage of analysis, the Total Requirements Statement, will elaborate the detailed, step-by-step processing to be performed and the detailed data models for the databases.

Warehouse Database

Contents:

- ❑ Soil data tables including all NASIS data elements, for all traditional (non-MLRA) soil surveys that have been approved for public release by State Soil Scientists.
- ❑ Spatial soil data, where available, for the same soil surveys.
- ❑ Other types of soil surveys, such as STATSGO or MLRA legends.
- ❑ Stored interpretation ratings generated from data and criteria in effect at the time a soil survey is approved for release.
- ❑ Prior versions of soils tables and maps, when superseded by a newer copy. Each version is identified by an effective date.
- ❑ Metadata including the contents of the NASIS data dictionary, spatial metadata, and other information needed to meet metadata requirements such as FGDC.
- ❑ Interpretation criteria, reports and queries similar to those used in NASIS.
- ❑ Supporting data including pedon descriptions, lab characterization data, performance measurements, field notes or images, as approved for public distribution by State Soil Scientists.
- ❑ Digitized soil survey area boundaries.
- ❑ Documents such as soil survey manuscripts, official series descriptions, soil taxonomy, FOTG tables, handbooks and manuals, when available in digital form.

Capabilities:

- ❑ Maintain linkage between soil features and tabular data.
- ❑ Support query and retrieval of data by either geographic or tabular attributes.
- ❑ Identify locations where tabular data or spatial data are not available to complete a national coverage, or where multiple maps exist for one area.
- ❑ Prevent modification of data other than by addition of a new version.
- ❑ Provide backup, mirroring, load balancing and other security measures to protect the integrity of the data.
- ❑ Permit conversion of the data structure from time to time to meet business requirements.

Capacity:

- ❑ A national set of tabular and spatial soil survey data will require about 100 GB of disk space. Adding space for other data, plus software, redundant storage, and work space yields an estimate of 300 to 500 GB for an initial configuration.
- ❑ The warehouse database is visualized as a data source for the Data Marts, rather than supporting direct access by users. This configuration is similar to the NASIS database server, so a system of similar performance capacity is projected.
- ❑ The warehouse database will be available around the clock, but downtime for scheduled maintenance of a few hours per week at non-peak hours is acceptable. Disaster recovery time of 1 to 2 weeks is acceptable.

Options:

- ❑ We recommend storing the NASIS calculated interpretation results, as well as the interpretation criteria, when a survey is placed in the warehouse. In principle the same interpretation results could be generated from the warehouse on demand, by using the data and criteria that were in effect at the time of survey certification. However, there are complications in ensuring that the correct versions of the interpretation criteria are in the warehouse when a survey version is certified. Since the intent is to produce the same results for a survey each time they are requested (until a new version is certified) and the amount of storage is not excessive, it makes sense to store the results.
- ❑ The Arc Spatial Database Engine (ArcSDE) from ESRI, Inc. is the recommended product for accessing the spatial data, since it can support a complete national coverage and linkage to the tabular data. No other products are comparable.
- ❑ The recommended database under ArcSDE is Informix, to facilitate the use of custom interpretation and reporting software originally written for NASIS. Use of other database software, such as Oracle or SQL Server, would require porting of the NASIS tools or development of new tools, and result in a delay of up to a year in implementing these core capabilities. The potential savings in initial software costs for a solution such as SQL Server are more than offset by the increased development costs.

The recent acquisition of Informix, Inc. by IBM casts doubts on the future of the Informix product. Because of the large Informix customer base, we are assuming that IBM will either continue to support Informix or provide a reasonable migration path. The NASIS database would also be affected by such a migration, and using the same architecture for both systems would reduce the overall cost of maintenance.

Staging Server

Contents:

- ❑ Transient storage of soil survey tabular and spatial data in a form similar to that in the warehouse database.
- ❑ Transient storage of query, report and interpretation criteria for testing.
- ❑ No permanent storage of data.

Capabilities:

- ❑ Accept certified soil survey data from NASIS and spatial data from a Digitizing Unit.
- ❑ Assign linkage (map unit key) to features and verify referential integrity with tabular data.
- ❑ Query, reporting and map display for data quality verification.
- ❑ Transfer data to warehouse database.
- ❑ Edit and test reports and interpretations (see Soil Reports and Interpretations Data Mart)
- ❑ Not open to the public. Access limited to authorized personnel.

Capacity:

- ❑ Storage of not more than 50 soil surveys at one time, plus associated data, requires 10 to 20 GB of disk space.
- ❑ The system should support 20 to 30 simultaneous users performing complex database and spatial analysis.
- ❑ Minimal usage is expected outside of normal working hours in North America.

Options:

- ❑ This represents a change in operating procedures for the Digitizing Units. Depending on the time required to design and implement the new procedures, the staging server might not be included in the first phase of warehouse development. Instead, certified surveys could be moved directly from NASIS to the warehouse, then exported for SSURGO certification in the same way they are now. The corresponding spatial data could be loaded into the warehouse from the SDE database currently maintained at NCGC. That would include surface point and line features not documented in NASIS.

FOTG / SSURGO Data Mart

Contents:

- ❑ Soil data tables, and associated digital spatial data where available, in SSURGO format.
- ❑ Current certified data for each soil survey as well as previously released versions as available.
- ❑ List of customers who receive data and wish to be notified of updates.
- ❑ Index of available FOTG data and documents.

Capabilities:

- ❑ Linkage to Natural Resources Data Gateway for locating available data sets.
- ❑ Web based access.
- ❑ Extract data from warehouse database, converting tabular and spatial data to SSURGO format.
- ❑ Download data for use in the Soil Data Viewer and other components of the Customer Service Toolkit.
- ❑ Download data to customers other than Service Centers, like the existing SSURGO distribution.
- ❑ Download tabular data without spatial data when digital spatial data is not available.
- ❑ Provide access to other documents from the warehouse as appropriate for the Field Office Technical Guide, Section 2, including digital manuscripts, images, etc.
- ❑ Send notices of data updates to registered customers.

Capacity:

- ❑ Archived data sets for the nation will require about 100 GB of disk space initially.
- ❑ There are currently about 200 SSURGO downloads from Fort Worth per day, which gives a lower bound for the capacity of this data mart.
- ❑ The data mart will be available around the clock, with scheduled downtime acceptable.

Options:

- ❑ In principle, a SSURGO package for a survey area could be extracted from the warehouse database on demand, but it is likely to be a resource intensive process. This data mart is designed to hold current and past SSURGO packages, as well as SSURGO formatted data tables where spatial data are not available, in an archived form for ready distribution.
- ❑ Some aspects of the FOTG need further analysis. For example, it is likely that data and documents in the warehouse will need to be available for online interaction in addition to downloading. The Web SDV Data Mart may be able to satisfy some of these needs. Another possibility to be investigated is automatic delivery (push) of updated FOTG data to the appropriate service centers rather than downloading on request.
- ❑ The FGDC standard for soils data needs to be updated, and it is hoped that it will be close to SSURGO in content. If so this data mart should be capable of delivering an export in FGDC format.

Web Soil Data Viewer

Contents:

- ❑ Current versions of soil surveys that have certified tabular and spatial data.
- ❑ A geospatial database separate from the warehouse database, using the architecture that supports the Web SDV, including SQL Server, ArcIMS, and Microsoft IIS.

Capabilities:

- ❑ Extract latest version of soil surveys from warehouse database and convert to appropriate format for Web SDV.
- ❑ Host the Web SDV application.

Capacity:

- ❑ The USDA Lighthouse project recently collected performance data that can be used to estimate processor requirements for this data mart. The number of simultaneous users is hard to estimate, and can be expected to grow over time as the service becomes known. Scalability of the system will be a major design goal.
- ❑ A complete national database for this application will require about 100 GB of disk space.
- ❑ The data mart will be available around the clock, with scheduled downtime acceptable.

Options:

- ❑ Eventually the Web SDV program could be converted to use the API described below. This data mart is intended to take immediate advantage of the prototype Web SDV.
- ❑ The geodatabase in this data mart could be made available for online geographic analysis to users having appropriate client software, such as ArcView.

Soil Reports and Interpretations Data Mart

Contents:

- ❑ Web based interface for selecting an area of interest (by legend, map unit, or arbitrary boundary) and type of output.
- ❑ Prewritten report scripts and interpretation criteria.

Capabilities:

- ❑ Generate interpretations and reports using query, report and interpretation scripts stored in the warehouse database. This includes applying new interpretation criteria to existing data.
- ❑ Deliver report output as HTML page, Postscript print file, or plain text file.
- ❑ Interface with GIS for spatial querying or map generation.
- ❑ Access any version of data in the warehouse database with appropriate query.

- ❑ Report on changes between versions.

Capacity:

- ❑ Processor load on this data mart is likely to be similar to that on the NASIS application servers. The number of simultaneous users will grow over time, so scalability of the configuration will be an important requirement.
- ❑ Disk storage is needed only for software and working space, about 10 GB per server.
- ❑ The data mart will be available around the clock, with scheduled downtime acceptable.

Options:

- ❑ This is intended to make use of existing NASIS tools for report and interpretation generation, to provide quick implementation. This does not preclude development or purchase of other tools at a later time.
- ❑ Because of the requirement to maintain past versions of data in the warehouse database, its structure will differ from that of NASIS, particularly in the key columns. Although some kinds of key changes can be supported transparently by NASIS, there may be some reports or interpretations that cannot run unchanged on both NASIS and the warehouse. Since modification of data, including report and interpretation scripts, is not permitted in the warehouse, a different environment is needed for testing and modifying these scripts. It is likely that the Staging Server can support this need.
- ❑ A set of reports to identify changes between versions of data in the warehouse is required. This cannot be addressed with NASIS style reports, so a special report program would have to be developed.

Custom Export Data Mart**Contents:**

- ❑ Data dictionary driven web interface for custom exports.
- ❑ User saved export configurations and standard export configurations.

Capabilities:

- ❑ Identify a set of map units and a set of attributes to produce a table or tables of data to be downloaded.
- ❑ Provide common functions for collapsing data into simpler structures, such as dominant component, surface layer, first restriction, representative value, top level interpretation, etc.
- ❑ Selection of output formats to match input requirements of various tools or models.
- ❑ Selection of metadata content and format.
- ❑ Delivery of supporting data (Pedons, lab data, documents, etc.).

Capacity:

- ❑ The processing and storage requirements for this data mart are similar to those for reporting and interpretations. It should be possible to share servers, with the option of adding servers and load balancing as the traffic requires.
- ❑ There are currently about 40 downloads per day from the ISU site, primarily MUIR and OSD files.

Options:

- ❑ This is intended to replace and extend the capabilities of the current MUIR download. An initial implementation could be to just reproduce the output options provided in the current MUIR site.
- ❑ This is an opportunity for incremental development, beginning with a small set of standard export options, and gradually adding more capabilities.

Soils Application Programming Interface (API)**Contents:**

- ❑ A documented set of software interfaces for retrieving data from the warehouse database for use in models or third party software packages.

Capabilities:

- ❑ Allow application programs to access the soils data online while preserving all security requirements and business rules.
- ❑ Insulate applications from some of the internal complexity of the warehouse database. For example, the API could provide access only to the current version of any soil survey, or present a view that matches an older database structure for a period of time after a structure change takes place.
- ❑ Allow different views for authorized users and the general public.

Capacity:

- ❑ The processing and storage requirements for this data mart are similar to those for reporting and interpretations. It should be possible to share servers, with the option of adding servers as the traffic requires.

Options:

- ❑ One approach would provide ODBC access to a database consisting of views into the warehouse database. This is relatively simple to set up, but requires the application programmers to do most of the work of navigating through the soils tables. Some business rules could not be enforced with this approach.
- ❑ A more extensive API could provide a number of capabilities to make the database easier to use, such as aggregating repeating group data, selecting representative values, identifying dominant components, and so forth. Enterprise Java Beans are a possible implementation technique.